
Subject: Re: Weighting district-level data

Posted by amira.elshal.1@city.ac.uk on Fri, 18 Dec 2015 19:43:59 GMT

[View Forum Message](#) <> [Reply to Message](#)

1) When I try to re-calculate the district-by-survey-round indicators again using the weights, I get the same results for each district (as in the unweighted case). I think this is probably due to the way I create/calculate the district-level indicator (below):

```
collapse (sum) women_using_modern_methods (count) women_interviewed, by(District)
gen %women_currently_using_modern_methods=(sum_of_women_using_modern_methods/count_of_women_interviewed)*100
```

Note: I denote women_using_modern_methods by 1 before calculating the district-level indicator.

When I use the weights as you have kindly suggested (below), the district-level indicators are the same as in the unweighted case:

```
gen myweight=v005/1000000
collapse (sum) women_using_modern_methods (count) women_interviewed [pweight=myweight],
by(District)
gen %women_currently_using_modern_methods=(sum_of_women_using_modern_methods/count_of_women_interviewed)*100
```

I do NOT simply create the district-level indicators as means of observations of individuals in these districts.

2) You made this point quite clear. Thank you.

3) I fail to do the clustering when I run the diff-in-diff regression below:

```
diff district_indicator, t (treated) p (t), cluster(District)
```

Stata 12 gives me the return code 198 of invalid syntax. Shall the 'District' variable (or whichever variable I use to cluster) be written down in a particular way?

4) The PSU is not the district in my case. I spatially join the displaced cluster locations of women to the GIS polygon data of Egypt's district boundaries. i.e. I allocate each woman to the relevant district.